KNOWLEDGE SECTORS FOR LOGICAL PROCESSING OF PATIENT DATA IN THE HELP SYSTEM

Homer R. Warner

Department of Medical Biophysics and Computing
College of Medicine
University of Utah

HELP is a system developed and implemented at the LDS Hospital in Salt Lake City, Utah to facilitate medical decision-making by physicians and other personnel who participate in the delivery of health care at this institution. It consists of a set of programs which accept data from more than 100 ports around the hospital, manage a patient-oriented file in which this data is stored, and from medical knowledge stored in the form of decisions criteria in another file, make decisions based on that data for every patient in the hospital, and distribute those decisions to the appropriate people. More than 190,000 patients have been served by the system as it has evolved both in scope and sophistication over the last twelve years.

There are several concepts that have dominated the design of HELP. Some of these differ considerably from the philosophy of the PROMIS system just described by Dr. Weed. First, the physician is not the primary source of data in the HELP system, but he is the primary recipient of information provided by the system as the result of interactions of the knowledge base with the patient's data. This information appears as parts of laboratory reports, is carried to the ward in some cases by specialized personnel such as the pharmacist or special nurses who follow-up on alerts generated by the HELP system on a hospital-wide basis, or is displayed on request from any terminal throughout the hospital. The second design concept is that each new piece of information added to a patient's file be enough to make a decision if considered in conjunction with other information already in that file, which is specified in the decision criteria. Thus, the system does not require that each user sit at a terminal and enter information at all in order to receive appropriate consultation on his patient. The system is data driven, that is, as each new item of information is added to any's patients file, the HELP system automatically loads into the CPU any knowledge sectors which use that item of information. Each of these sectors is processed using this new piece of information, as well as other information in the patient's file which is specified by the decision criteria (the HELP sector). If one or more decision criteria are satisfied, those decisions are stored on the patient's file.

A third criterion adopted in the design of the HELP system is that the system be modular in order to permit implementation in a form that would be useful even with a small subsets of the eventual knowledge base that the system will acquire. This modularity also provides two other important benefits. It greatly facilitates maintenance of the system as the medical ware becomes more complex and inclusive and it provides a workable schema for assigning direct responsibility for each decision the system makes a specific medical person. The 300 practicing physicians on our staff who are the primary recipients of the HELP consultations are instructed to consider decisions made by the system as consultations from the mediclal expert responsible for the particular block of HELP sectors corresponding to the subject matter area of the consultation. HELP is not an inanimate big brother looking over your shoulder but is simply a means of providing a consultation with the expert in question where appropriate on a twenty-four hour a day basis.

HELP provides a convenient system for this medial expert not only to create a HELP sector on an interactive mode using natural language and a sequence of prompts from HELP, but it also provides useful tools for him to debug his logic, test it for a period of time before he calls for the decisions to start going to the attending physician and, most importantly, provide him with tools that will let him make use of the data base to create new and improved old decisions criteria for future use. The importance of providing for such self-improvement in the system can be appreciated by those of you who have attempted to explicitly state the logical basis for a decision only to recognize how flimsy is the base upon which our criteria rests. With current medical knowledge as a starting point, an algorithm can often be developed which is useful and will serve until enough experience is accumulated to provide the basis for improvement.

To illustrate the process by which decision criteria are derived from the data base, an example will be presented which involves the use of STRATO. This is a program which allows the user to create subpopulations of the data base as a list of patient numbers for patients who satify criteria defined by a HELP sector. The investigator writes the HELP sector by interacting through a terminal with the HELP compiler to specify the variables upon which the decision to include a patient will be based, as well as the temporal and logical or probabilistic relationships among these variables. The data from each patient in the population specified is analyzed using these criteria and the population is given a label. A variable may be created in the same way by writing a HELP sector which defines the data item to be retrieved from each patient and stored as a list with the variable name and the population from which it was derived in the same order as the patients in the population list. The first slide shows a list of the STRATO options and the next slide shows an example of a sequence of interactions between STRATO and the user as he prepares to compare two variables by generating histograms to represent their distribution. In this case, the variable is the first serum calcium value drawn during the time interval from two hours before admission to the hospital until twenty-four hours after admission. In patients who were admitted to the coronary care unit, but who did not receive drugs for the treatment of arrhythmia during their stay, the test variable is the first serum calcium value over the same time interval in a population of patients who did receive such drugs. The question asked is, "Can the serum calcium value on admission provide information that would be useful in determining whether a patient is likely to really use the services provided by that unit?" The next slide shows the two histograms superimposed. The C's represent the control group and the T's the test group. The patient who developed arrhythmia, the test group, indeed have slightly lower serum calcium values on the average. The non-parametric Mann Whitney test was chosen for the analysis and showed the two populations to be different in terms of this variable at the two percent level. On the next slide the same data is displayed in another histogram in which each bin has approximately the same number of patients in it and reflects only the relative frequency from the two populations. This kind of representations has the advantage for use in decision-making in that each bin has approximately the same reliability and the more densely populated parts of the spectrum of values are more finely divided to provide additional accuracy. The next slide shows the list of percent of each population falling into each bin in the last two columns. Such data can be inserted directly into a HELP sector for use as an element in a Bayesean decision scheme. The next slide illustrates the performance of a simple decision sector created from this distribution. The histogram displays the estimate of probability that each patient in the two populations, that is those who received treatment, the test group, and those who did not, the control group, indeed belong to the test group. Obvioulsy, other variables must be included to make this decision with sufficient reliability to be useful, but this simple example will serve to illustrate how such medical knowledge can be acquired with a system such as HELP.

Now let us look in more detail at the sophistication needed to realistically represent medical knowledge and be of assistance in the everyday practice of medicine. The next slide is a HELP sector designed to select patients for coronary arteriography. Each sector begins with a messege which is to be distributed through one or more of the roots described earlier, if the final evaluation statement is true. The terms "F" and "G" refer to items labeled in order alphabetically in the body of the HELP sector. Each of these items may be a request for certain data in the patient's file or the specification of logical arithmetic or probabilistic functions to be performed on prior items in the list. The first item, labeled "A", requests a search for a record indicating that angina pectoris was diagnosed from history data by an other sector. This demonstrates that a decision made by a HELP sector is stored like any other data in the patient record and may be used itself for making another decision. In this example, items "B" and "C" are also prior decisions. Item "C" contains two terms followed by a Boolean statement "A or B" which is performed during a search and, if true, item "C" becomes true. Thus operations can be performed involving Boolean or arithmetic manipulations of items within a particular data field. Item "D" is an exist statement, that is , if "A" was not found, that is, the patient did not have angina pectories, then "E" is set to 80 and processing of the sector continues at item "F". Item "E" illustrates the use of probabilities derived from a distribution which is the one just dicussed. In this case we are looking at the distribution of probabilities that a patient has angina pectoris. To derive these probabilities, we examined two populations, one where patients who had coronary arteriography which demonstrated at least seventy-five percent obstruction of a coronary artery, and the other group were patients who had the test but showed insufficient obstruction to justify surgery. The angina pectoris sector generates a probability value, distribution of which is represented by this set of eight bins separated by comas. The first number in each set is the minimum value of the bin, for instance, 400 followed by 6 and 17, which represents the relatively likelihood of a patient from the test and control group having a probability of angina pectoris between 400 and 505. These probabilities are expressed per thousand. The first number 80 represents the a priori probabilitiy of the sector being true in the population of people on which it is run. That population is made up of people who have one or more of the data items used by this sector, since the sector is data driven. In this case the sector will be called if angina pectoris from history exceeds a probalility of .4 (the threshold set on this sector), an abnormal exercise ECG is recognized by HELP, or the HELP sector classifies a resting ECG as an old inferior or an old anterior infarction. The a priori probability then is 180 since 18 percent of the patients on whom this sector is called have coronary arteriography which proves positive, as defined above.

Item "F" uses the probability calculated in item "E" as the new a priori probability and then based on the presence or absence of item "B" the positive exercise test calculates a new probability using two values. The first, 710 represents the probability of item "B", the abnormal exercise test in a patient with a positive coronary arteriography, and the second number, 345, is the probability per thousand of a positive exercise test in patients who are considered for coronary arteriography but either don't have the tests done or if they do the test for significant coronary occlusion is negative. This process continues then with item "G". Here the probability of an old infarct by ECG is fairly unlikely in both populations because the ratio 160 to 55 is relatively large, the presence of this pattern will have a big influence on the probability of this sector beeing true, but the absence of the pattern will have relatively little effect since the complement (1,000 - these numbers) result in a relatively small ratio (840/945). The final evaluation statement for this sector is a test of whether the final probability "G" is greater than a threshold probability "H" which was calculated by another sector. The principles involved in this calculation will be described in some detail, since the estimation of this threshold probability plays a critical role in decision-making. Next slide please. Before doing that, however, I would like to discuss briefly the sector for diagnosing angina pectoris from a self-administered history. The set of questions which are answered "yes" more frequently or less frequently by patients who subsequently are discharged with the diagnosis of angina than in patients without angina are chosen to perform this decision. Skipping down to item "J" we see age represented as a distribution in the two sets of patients. Questions such as "A" are represented by two bits in the patient's record. The first bit indicates whether the patient was asked the question, and the second bit reflects his answer. If both bits are set to zero, the term takes on the a priori probability and the question is ignored. If the answer is "no" this, too, represents information and the complements of the two probabilities are used in the Bayesean calculation. The next slide shows a continuation of this sector. The last two statements make use of threshold values to exit and thus label the sector false; otherwise, the final probability is stored in the patient's record. The next slide shows the method used by HELP to estimate the probability threshold for a decision. Consider this 2 x 2 matrix presenting the costs of treating or not treating a patient who has or does not have a particular diagnosis. The term "A", then, is the cost of treating the patient if the patient proves to actually have the diagnosis, while "C" is the cost of not treating the patient if indeed he has that diagnosis. The cost of treating the patient if the diagnosis is uncertain is equal to P, the probability that he has the diagnosis, x A + 1-P x B, the cost of treating him if he doesn't have the diagnosis. Likewise the cost of not treating him is PC + 1 - P x D. Now the value of P, which causes these two terms to be equal, represents the threshold of probability above which it costs less to treat the patient than to not treat him, and can be determined by setting these two expressions equal to one another and solving for P. This expression is valid as long as D does not equal B and the calculated value of P is in the range of 0 thru 1. Next slide please. A HELP sector can be used to estimate each of the four terms, A, B, C, and D needed to calculate the threshold using the following representation. For each term we must use this expression involving the following terms: LE is the life expectancy under this option independent of any immediate mortality, such as death occurring during an operation, or before discharge from the hospital. This is used to estimate the number of good days the patient has left. A good day is a day in which the patient can operate at his capacity when he is in good health. To estimate the good days, the days lost due to ill health must be subtracted. This can be estimated for a particular term by summing the number of days D with a given level of lost capacity L and summing this product with the number of levels represented. The lost capacity is equal to 1 minus the ratio of the capacity at that level or the capacity in good health. A correction must be made for the immediate mortality M by multiplying the good days by 1-M, which is the probability that the patient will survive long enough to enjoy them. The expected good days are then multiplied by 70, a somewhat arbitrary number of dollars, representing the value of one good day of life based on what several social programs have been willing to pay to keep a person alive over an extended period. From this dollar value, then, is subtracted three elements of cost. The first is the immediate cost of treatment, such as surgery or some other one-time procedure; CC is a continuing cost such as a medication that must be taken for the rest of the patient's life; and LI is lost income, which is obtained from the product of the patient's salary and twice the number of lost days, since the patient probably will not go to work even when he can perform at up to half his capacity. The last slide illustrates the application of this algorithm to the calculation of probability threshold for a fifty-year-old male who is being considered for coronary arteriography. For example, term A is calculated by multiplying his life expectancy in years by 365 and subtracting the number of days lost if indeed the test is done and proves positive and he goes to surgery. Immediate mortality, including the surgery, is 1-1/2 percent and the days lost is calculated as shown on the second line. The first day post-operative is almost a total loss .9. For the next three days he can expect to function at 40 percent of capacity, meaning a loss of .6. The next ten days at half capacity, then for sixty days his loss is .15, and for the remainder of his life, that is, the life expectancy minus 74, the days already accounted for in the preceding terms, he will have a residual handicap of .05. The fixed immediate cost, IC, averages $9500, while his lost income is his annual wage of $16,000 divided by 365 times twice his lost days. Using the values calculated in a similar way for all four terms, the calculated probability threshold for deciding whether to perform the coronary angiogram on this patient is .61. Lights please.

At the present time over 1400 HELP sectors are operational at the LDS hospital covering such areas as history, ECG interpretation, CCU and ICU management, potential adverse drug reaction detection, blood gas and electrolyte recognition, prediction of x-ray interpretation to facilitate reporting, cardiovascular and pulmonary laboratory data interpretation, suggestions for follow-up laboratory procedures, outpatient disease management protocols, hospital-wide on-line quality of care assessment, management and analysis of clinical trials, and management of patients with end stage renal disease. The central decision-making and file management system is on a Control Data 3300 machine. Much of the data gathering and pre-processing is performed by minicomputers in the clinical laboratory, the catheterization laboratory, the pulmonary function lab, the general medical-surgical intensive care unit, the coronary care unit, and admission screening laboratory. Plans are underway to convert the central system to new hardware using a high level language, so that it can be more easily implemented at other institutions. The HELP secors themselves are expressed in natural language and can be readily transfered. Our experience with HELP continues to strengthen the conviction which motivated its development; namely, that the most useful role of computers in medicine will be in uniting a knowledge base with a patient data base to provide expert assistance to the physician in decision-making.

Send inquiries to:

Homer R. Warner, M.D., Ph.D.
Professor and Chairman
Department of Biophysics and Computing
LDS Hospital
325 Eighth Avenue
Salt Lake City, UT 84143